Single Nucleotide Polymorphisms and Body Mass Index: The Framingham Heart Study

Jing Hua Zhao[1]*, Jian'an Luan[1], Qihua Tan[2], Shengxu Li[1], Fuzhong Xue[3], Wendi Qian[4], Ruth JF Loos[1], Nicholas J Wareham[1]

[1]MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK
[2]Department of Biochemistry, Pharmacology and Genetics, Odense University Hospital, Sdr Boulevard 29, Odense C., DK-5000, Denmark
[3]Institute of Public Health, Shandong University, 44 Wen Hua Xi Lu, Jinan 250012, PR China
[4]MRC Clinical Trials Unit, 222 Euston Road, London, NW1 2DA, UK

*Corresponding author

Email addresses:
JHZ: jinghua.zhao@mrc-epid.cam.ac.uk
JAL: jianan.luan@mrc-epid.cam.ac.uk
QT: qtan@health.sdu.dk
SL: shengxu.li@mrc-epid.cam.ac.uk
FX: xuefzh@sdu.edu.cn
WQ: wq@ctu.mrc.ac.uk
RJFL : ruth.loos@mrc-epid.cam.ac.uk
NJW : nick.wareham@mrc-epid.cam.ac.uk

Abstract

The recent rise in obesity may be attributable to changes in the environment yet genetic predisposition is likely to account for substantial individual differences. As a key variable to define obesity, body mass index (BMI, weight (kg)/height $(m)^2$) is of primary interest but a good understanding of its genetic and environment determinants has been hampered by the fact that the genetic variants were largely unknown. This is expected to change with the recently published single-nucleotide polymorphisms (SNPs) through genome-wide association studies (GWASs), and therefore to verify these findings in large epidemiological cohorts with comprehensive measurements is an important step for further investigation. We here conducted a genome-wide association analysis of 6848 unrelated and related individuals in the Framingham Heart Study with focus on some recently established genetic variants and BMI. We showed that single-nucleotide polymorphisms (SNPs) nearby *FTO*, *TMEM18*, *MC4R* and moderately *INSIG2* genes have significant association with BMI. Our analysis can be seen as the first step towards further investigation into these findings.

Background

It was estimated that in 2005 approximately 1.6 billion adults (age 15+) were overweight and at least 400 million adults were obese[1]. The rising prevalence of obesity is associated with an increase in diabetes, coronary heart disease, breast and colon and endometrial cancers, mortality[2], and elevated economic cost[3]. The recent identification of obesity-susceptibility loci[4-7] provides opportunity for an improved understanding of genetic and environmental contribution[8,9] in the etiology[10] and a great way forward to developing strategy for prevention and treatment. It is important to verify these findings in large studies with further data.

Body mass index (BMI) defined as weight in kilograms divided by the square of height in meters (kg/m^2) is a useful measure of relative obesity such that a normal range for BMI is 18.5~24.9 kg/m^2, an overweight range is 25.0~29.9 kg/m^2, and obesity is $\geq$ 30 kg/m^2[11]. We here conduct a genome-wide association analysis of BMI in the Framingham Heart Study (FHS) [12], a distinguished cohort consisting of both population-based and family-based samples and comprehensive measures of genetic variants on Affymetrix 500K GeneChip, with particular focus on SNPs nearby *FTO* (rs1121980, rs9939609), *TMEM18* (rs6548238), *MC4R* (rs17782313, rs17700633) and *INSIG2* (rs7566605) genes and rs1106683 from FHS[13,14], and *NYD-SP18* (rs745229) from Wilk et al.[15] who reported significant association of with BMI but not *FTO*.

Methods

The FHS is under the direction of National Heart, Lung, and Blood Institute (NHLBI) which began in 1948 with the recruitment of adults from the town of Framingham, Massachusetts. Data available for the Genetic Analysis workshop 16 were 7130 individuals from the original cohort (373), the first generation cohort (2760) and the third generation cohort (3997) with sex, age, height, weight, blood pressure, lipids, smoking and drinking. There were 6848 (3118 males and 3730 females) individuals, aged between 5 and 72 at baseline, with genomic data based on Affymetrix 500K GeneChips, among whom 227 were unrelated and the remaining were relatives from 962 families. As most individuals had BMI measured at baseline (6746); SNP-BMI

association was assessed using baseline BMI and additive model for autosomal SNPs with adjustment for age and sex, excluding individuals under age of 18 for their relatively small number (89). Individuals eligible for the analysis had mean (SD) age of 37.7(9.0) and BMI of 26.0 (5.0), of which 3390 were overweight and 1123 obese. Further exclusion was made via PLINK[16] v1.05 for individuals with genotyping rate <=0.9 (4) and SNPs with Hardy-Weinberg equilibrium test p_{HWE}<0.000001 (10541), failure of missingness test (9882), and minor allele frequency (MAF)<0.01 (62232), leading to 408648 in a total of 488146 SNPs in the analysis. While genome-wide association results were obtained, we focused on the recently identified genetic variants as described earlier and we made no attempt to correct for multiple testing. As SNPs (in or near genes) rs6235 (*PCSK1*), rs10938397 (*GNPDA2*), rs7498665 (*SH2B1*), rs10838738 (*MTCH2*), rs11084753 (*KCTD15*), rs2815752 (*NEGR1*) are not on the Affymetrix 500K GeneChip, genotype imputation were done ~50kb flanking either side of the candidate SNPs for all individuals in the analysis based on HapMap CEU sample via IMPUTE[17] v0.5 and the analysis of imputed genotypes were done with SNPTEST[17] v.1.5 according to National Center for Biotechonology Information (NCBI) Build 36. The Q-Q and Manhattan plots were produced with R/gap[18] v1.0-18. All analyses were done on our Linux clusters.

Results

The association test p values deviated from but appeared to conform to the expected uniform distribution after correcting for the association statistics with an inflation factor of 1.41 (Figure 1) as estimated from R/snpMatrix[19] v1.6.1. To investigate if this was due to inclusion of unrelated individuals and families with single member (197), results were also obtained from families with at least two members and the results were quite similar (not shown). The basic information for SNPs and results of SNP-BMI association are shown in Tables 1 and 2. SNPs nearby *FTO*, *TMEM18* and *MC4R* were statistically significant at level of 0.05 and rs7566605 as reported earlier in the same cohort was marginally significant. It is notable that the two SNPs near *MC4R* showed low linkage disequilibrium in this data (r^2=0.15) and the HapMap CEU sample (r^2=0.11), suggesting both are important in the regulation of body mass. Genotype imputation based on HapMap for the non-Affymetrix 500K GeneChip SNPs yielded no significance at the 0.05 level (all p>0.35). To confirm these findings, analyses were further conducted with inverse normal transformation (as in SAS) of residuals after regressing BMI on sex and age while accounting for familial relationship, i.e., genome-wide rapid association using mixed model and regression (GRAMMAR)[20] using R/GenABEL[21] v1.4-2. This has given similar results.

Discussion

Our analysis of the Framingham study showed SNP-BMI association involving *FTO*, *TMEM18*, *MC4R* and moderately *INSIG2* genes. Difference in degree of statistical significance between family-based association test and generalized estimating equation as seen earlier[14] regarding association of rs110683 and rs7566605 with mean BMI in this cohort seemed to support our finding with respect to rs110683. Our findings can potentially be strengthened with formal checks for population stratification, genotyping errors, missing data, misspecification of familial relationship and cross-validation via sample partition. Our genome-wide results may serve as a resource for SNP-BMI associations perhaps in light of additional information from the literature, replication from further samples from this cohort and

meta-analysis with other study samples. The precision of estimate of per-allele change in BMI based could then be improved with larger sample sizes[5,7,22]. A compelling extension of current analysis is the modeling of the interrelationship between multiple traits and multiple genes which goes beyond gene-discovery for gaining insight into complex traits such as BMI. Our earlier work suggested that one can have considerable insight into differential effects of SNPs on BMI-related anthropometric measurements as in[4,7,23] via structural equation modeling. The use of multiple variants such as those nearby *FTO*, *TMEM18* and *MC4R* genes would be also appealing, and estimation of their joint genetic relative risk can be facilitated through obesity outcomes and family-based association test[7] [24]. As a further step beyond consideration of mean BMI[14], it will be of considerable interest to study candidate SNPs on change of BMI and obesity-disease association in light of other disease-predisposing variants[25] and pathways[7,26,27].

Conclusions

Using baseline measure of body mass index from three generations of the Framingham cohort, we showed SNP-BMI association involving *FTO*, *TMEM18* and *MC4R*, and moderately *INSIG2* genes, further work in light of these findings will be of considerable interest.

Acknowledgements

The Genetic Analysis Workshops are supported by NIH grant R01 GM031575 from the National Institute of General Medical Sciences. We thank NIH, the Framingham study and GAW16 for enabling us to access the data and giving insight into the analysis. We are grateful of colleagues within Medical Research Council for administrative and computing support and in the Genetic Investigation in Anthropometric Traits (GIANT) consortium for collaborative work. Our final thanks are to Dr Jemma B. Wilk for clarification regarding Family Heart Study and Dr Eleftheria Zeggini regarding genotype imputation with family data.

References

1. World Health Organization. Obesity and overweight. 2006.

2. Flegal KM, Graubard BI, Williamson DF, Gail MH: **Cause-specific excess deaths associated with underweight, overweight, and obesity.** *JAMA* 2007, **298:** 2028-2037

3. Foresight. Tackling Obesities: Future Choices -- Summary of Key Messages. 2007. Govenment Office for Science 2007.

4. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW *et al.*: **A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.** *Science* 2007, **316:** 889-894

5. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, Inouye M, Freathy RM, Attwood AP, Beckmann JS *et al.*: **Common variants near MC4R are associated with fat mass, weight and risk of obesity.** *Nat Genet* 2008, **40:** 768-775

6. Benzinou M, Creemers JW, Choquet H, Lobbens S, Dina C, Durand E, Guerardel A, Boutin P, Jouret B, Heude B *et al.*: **Common nonsynonymous variants in PCSK1 confer risk of obesity.** *Nat Genet* 2008, **40:** 943-945

7. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, Berndt SI, Elliott AL, Jackson AU, Lamina C *et al.*: **Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.** *Nat Genet* 2009, **41:** 25-34

8. Maes HH, Neale MC, Eaves LJ: **Genetic and environmental factors in relative body weight and human adiposity.** *Behav Genet* 1997, **27:** 325-351

9. Wardle J, Carnell S, Haworth CM, Plomin R: **Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment.** *Am J Clin Nutr* 2008, **87:** 398-404

10. Li S, Loos RJF: **Progress in the genetics of common obesity: size matters.** *Current Opinion in Lipidology* 2008, **19:** 121

11. Khan MG: *Encyclopedia of Heart Diseases*. Elsevier 2006

12. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Sr., Fox CS, Larson MG, Murabito JM *et al.*: **The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination.** *Am J Epidemiol* 2007, **165:** 1328-1335

13. Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T, Wichmann HE, Meitinger T, Hunter D, Hu FB *et al.*: **A common genetic variant is associated with adult and childhood obesity.** *Science* 2006, **312:** 279-283

14. Fox CS, Heard-Costa N, Cupples LA, Dupuis J, Vasan RS, Atwood LD: **Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project.** *BMC Med Genet* 2007, **8 Suppl 1:** S18

15. Wilk JB, Laramie JM, Latourelle JC, Williamson S, Nagle MW, Tobin JE, Foster CL, Eckfeldt JH, Province MA, Borecki IB *et al.*: **NYD-SP18 is associated with obesity in the NHLBI Family Heart Study.** *Int J Obes (Lond)* 2008, **32:** 930-935

16. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ *et al.*: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81:** 559-575

17. Marchini J, Howie B, Myers S, McVean G, Donnelly P: **A new multipoint method for genome-wide association studies by imputation of genotypes.** *Nat Genet* 2007, **39:** 906-913

18. Zhao JH: **gap: genetic analysis package.** *J Stat Soft* 2007, **23:** 1-18

19. Clayton D, Leung HT: **An R package for analysis of whole-genome association studies.** *Hum Hered* 2007, **64:** 45-51

20. Aulchenko YS, de Koning DJ, Haley C: **Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis.** *Genetics* 2007, **177:** 577-585

21. Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM: **GenABEL: an R library for genome-wide association analysis.** *Bioinformatics* 2007, **23:** 1294-1296

22. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, Styrkarsdottir U, Gretarsdottir S, Thorlacius S, Jonsdottir I *et al.*: **Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity.** *Nat Genet* 2009, **41:** 18-24

23. Freathy RM, Timpson NJ, Lawlor DA, Pouta A, Ben-Shlomo Y, Ruokonen A, Ebrahim S, Shields B, Zeggini E, Weedon MN *et al.*: **Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI.** *Diabetes* 2008, **57:** 1419-1426

24. Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ: **Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer.** *N Engl J Med* 2008, **358:** 2796-2803

25. Kenchaiah S, Evans JC, Levy D, Wilson PW, Benjamin EJ, Larson MG, Kannel WB, Vasan RS: **Obesity and the risk of heart failure.** *N Engl J Med* 2002, **347:** 305-313

26. Dong LM, Potter JD, White E, Ulrich CM, Cardon LR, Peters U: **Genetic susceptibility to cancer: the role of polymorphisms in candidate genes.** *JAMA* 2008, **299:** 2423-2436

27. Zhao JH, Luan JA, Baksh F, Tan Q: **Mining gene networks with application to GAW15 problem 1.** *BMC Proceedings* 2007, **1 (Suppl 1):** S52

Tables

Table 1 - Characteristics of selected SNPs (the first allele is minor, N=6259)

SNP	Nearby gene	Chr	Position (bp)	MAF	Alleles	p_{HWE}
rs6548238	*TMEM18*	2	624905	0.17	TC	0.96e-3
rs7566605	*INSIG2*	2	118552495	0.33	CG	0.18
rs745229	*NYD-SP18*	7	128146127	0.22	AC	0.28
rs1106683		7	131104065	0.13	AG	0.03
rs1121980	*FTO*	16	52366748	0.44	AG	0.57
rs9939609	*FTO*	16	52378028	0.40	AT	0.19
rs17782313	*MC4R*	18	56002077	0.22	CT	0.11
rs17700633	*MC4R*	18	56080412	0.31	AG	0.42

Table 2 - Mean (SD) of BMI (kg/m^2) by homozygote minor, heterozygote and homozygote major and association with selected SNPs

SNP	Mean (SD) (kg/m^2)			*β (SE)*	*P*	P_{GRAMMA}
rs6548238	25.56 (4.38)	25.66 (4.90)	26.08 (5.07)	-0.44 (0.12)	0.0012	0.0013
rs7566605	26.32 (5.06)	25.79 (5.05)	25.88 (4.99)	0.20 (0.09)	0.054	0.030
rs745229	25.27 (4.97)	26.03 (5.16)	25.88 (4.96)	-0.052 (0.10)	0.66	0.64
rs1106683	26.46 (4.69)	26.05 (5.09)	25.84 (5.04)	0.22 (0.12)	0.14	0.12
rs1121980	26.45 (5.27)	25.86 (5.00)	25.63 (4.92)	0.35 (0.08)	3.8e-4	6.2e-5
rs9939609	26.58 (5.19)	25.88 (5.01)	25.62 (4.94)	0.40 (0.08)	6.8e-5	5.9e-6
rs17782313	26.34 (5.71)	26.09 (5.16)	25.76 (4.90)	0.31 (0.10)	0.0094	0.028
rs17700633	26.31 (5.28)	26.00 (5.22)	25.76 (4.84)	0.27 (0.09)	0.013	0.038

Figures

Figure 1 - Q-Q plot (top panel) and Manhattan plot (bottom panel)